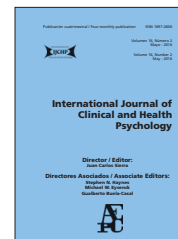




International Journal of Clinical and Health Psychology

www.elsevier.es/ijchp



ARTICULO ORIGINAL

Criteria-Based Content Analysis (CBCA) reality criteria in adults: A meta-analytic review

Bárbara G. Amado^a, Ramón Arce^{a,*}, Francisca Fariña^b y Manuel Vilariño^a

^a *Universidade de Santiago de Compostela, España*

^b *Universidade de Vigo, España*

Recibido 24 de noviembre de 2015; aceptado 20 de enero de 2016

PALABRAS CLAVE

Criteria-Based Content
Analysis;
adultos;
declaraciones;
credibilidad;
meta-análisis.

Criterios de realidad del Criteria-Based Content Analysis (CBCA) en adultos: una revisión meta-analítica

Resumen

Antecedentes/Objetivo: El *Criteria-Based Content Analysis* (CBCA) constituye la herramienta mundialmente más utilizada para la evaluación de la credibilidad del testimonio. Originalmente fue creado para testimonios de menores víctimas de abuso sexual, gozando de amparo científico. Sin embargo, se ha generalizado su práctica a poblaciones de adultos y otros contextos sin un aval de la literatura para tal generalización. **Método:** Nos planteamos una revisión meta-analítica con el objetivo de contrastar la Hipótesis Undeutsch y los criterios de realidad del CBCA para conocer su potencial capacidad discriminativa entre memorias de eventos auto-experimentados y fabricados en adultos. **Resultados:** Los resultados confirman la hipótesis Undeutsch y validan el CBCA como técnica. No obstante, los resultados no son generalizables y los criterios *auto-desaprobación* y *perdón al autor del delito* no discriminan entre ambas memorias. Además, se encontró que la técnica puede ser complementada con criterios adicionales de realidad. El estudio de moderadores mostró que la eficacia discriminativa era significativamente superior en estudios de campo en casos de violencia sexual y de género. **Conclusiones:** Se discute la utilidad, así como las limitaciones y condiciones para la transferencia de estos resultados a la práctica forense.

© 2015 Asociación Española de Psicología Conductual. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

*Correspondencia: Departamento de Psicología Organizacional, Jurídico-Forense e Metodología das Ciencias do Comportamento, Facultade de Psicología, Campus Vida, s/n, 15782 Santiago de Compostela, España.

Correo electrónico: ramon.arce@usc.es (R. Arce)

KEYWORDS

Criteria-Based Content Analysis;
Adults;
Statements;
Credibility;
Meta-analysis.

Abstract

Background/Objective: Criteria-Based Content Analysis (CBCA) is the tool most extensively used worldwide for evaluating the veracity of a testimony. CBCA, initially designed for evaluating the testimonies of victims of child sexual abuse, has been empirically validated. Moreover, CBCA has been generalized to adult populations and other contexts though this generalization has not been endorsed by the scientific literature. **Method:** Thus, a meta-analysis was performed to assess the Undeutsch hypothesis and the CBCA checklist of criteria in discerning in adults between memories of self-experienced real-life events and fabricated or fictitious memories. **Results:** Though the results corroborated the Undeutsch Hypothesis, and CBCA as a valid technique, the results were not generalizable, and the *self-deprecation* and *pardonning the perpetrator* criteria failed to discriminate between both memories. The technique can be complemented with additional reality criteria. The study of moderators revealed discriminating efficacy was significantly higher in filed studies on sexual offences and intimate partner violence. **Conclusions:** The findings are discussed in terms of their implications as well as the limitations and conditions for applying these results to forensic settings.

© 2015 Asociación Española de Psicología Conductual. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

La credibilidad del testimonio, principalmente del denunciante y especialmente en los delitos cometidos en el ámbito privado (e.g., sexuales, violencia familiar), es el elemento clave en la formación de juicios legales (Novo y Seijo, 2010), estimándose que afecta a nivel mundial aproximadamente al 85% de los casos (Hans y Vidmar, 1986). Si bien se han creado y contrastado numerosas herramientas para la evaluación de la credibilidad (Vrij, 2008), el Criteria-Based Content Analysis [CBCA] (Steller y Köhnken, 1989) es la técnica de referencia, gozando de alto respaldo científico (Amado, Arce y Fariña, 2015) y aceptación en la sala de justicia en diferentes países (Steller y Böhm, 2006; Vrij, 2008). Aunque la técnica fue originalmente formulada para ser aplicada al testimonio de menores víctimas de abuso sexual, se ha generalizado su uso a adultos, testigos, agresores y otras casuísticas en los Institutos de Medicina Legal ante la demanda judicial (Arce y Fariña, 2012). La revisión meta-analítica de Amado et al. (2015) encontró que la técnica, la cual se fundamenta en la Hipótesis Undeutsch (1967) que establece que las memorias de hechos vividos o auto-experimentados difieren en contenido y calidad de las memorias de hechos fabricados o inventados, era igualmente válida en otros contextos y rango de edades hasta los 18 años. Previamente a esta revisión, la literatura ya había contrastado la validez de la hipótesis en población de adultos y en diferentes contextos (Vrij, 2005, 2008), al tiempo que, dado que la hipótesis se sustentaba en los contenidos de memoria, ya se había sostenido teóricamente que la hipótesis sería igualmente aplicable a adultos y contextos ajenos al abuso sexual (Berliner y Conte, 1993).

El CBCA se estructura en torno a 19 criterios de realidad (véase la Tabla 1) que se reagrupan en dos factores, cognitivo (criterios 1 a 13) y motivacional (criterios 14 a 18). Acorde a los autores del CBCA (Steller y Köhnken, 1989), ambos factores se sustentan en la hipótesis Undeutsch, pero Raskin, Esplin, y Horowitz (1991) advirtieron que sólo 14 criterios de realidad se sustentaban en la citada hipótesis.

Tabla 1 Criterios adicionales.

- Estilo de la declaración (extensa, el entrevistado describe aspectos irrelevantes sobre los que no se le preguntó).
- Dar muestras de inseguridad (incertidumbre en la descripción de un ítem).
- Dar explicaciones de la falta de memoria (da razones por las que no es capaz de dar una descripción detallada).
- Clichés (expresiones o intervenciones que introducen demoras en la declaración).
- Repeticiones (elementos ya descritos se repiten sin información adicional).

El CBCA ha sido ampliado con categorías adicionales, unas aplicables a todos los contextos (Tabla 1) (Höfer, Köhnken, Hanewinkel y Bruhn, 1993) y otras para casos específicos (Arce y Fariña, 2009; Juárez, Mateu y Sala, 2007; Volbert y Steller, 2014), así como combinado con otras técnicas con substratos teóricos diferentes como los atributos de memoria (Vrij, 2008).

Dado que por la vía de los hechos (práctica forense) se ha venido utilizando el CBCA como una herramienta para discriminar memorias de adultos de hechos auto-experimentados y fabricados en diferentes tipos de casos, al tiempo que la literatura es numerosa, confusa (e.g., diseños que no se ajustan a los requerimientos para la aplicación del CBCA, conclusiones de efectos no significativos no apoyados por los datos dado que la potencia estadística de los estudios era insuficiente, $1-B < 0,80$) y contradictoria sobre el uso del CBCA en adultos, nos planteamos una revisión meta-analítica con los objetivos de contrastar la hipótesis Undeutsch en población adulta, así como la eficacia discriminativa de los criterios de realidad del CBCA y adicionales; el efecto de contexto (tipo

de caso), el efecto del entrenamiento en el engaño, el estatus del declarante, y del paradigma de investigación.

Método

Búsqueda de literatura

Se llevó a cabo una búsqueda exhaustiva de la literatura científica para identificar trabajos empíricos en los que se aplicara análisis de contenido al testimonio de adultos con el objeto de discriminar entre declaraciones de hechos auto-experimentados y fabricados, tanto de hechos inventados deliberadamente como implantados. En la búsqueda de estudios se procedió con una aproximación multimétodo en meta-buscadore (Google, Google Scholar, Yahoo), bases de datos científicas de referencia (PsycInfo, MedLine, Web of Science, Dissertation Abstracts International), redes sociales académicas para el intercambio de conocimiento dentro de la comunidad científica (i.e., Researchgate, Academia.edu), *ancestry approach* (cotejo de las listas de referencias de los trabajos seleccionados), y contacto con investigadores solicitando estudios no publicados mencionados en otras publicaciones. Se creó un listado de descriptores por aproximaciones sucesivas (i.e., al que se fueron añadiendo descriptores de las palabras clave de los artículos seleccionados): *reality criteria, content analysis, verbal cues, verbal indicators, testimony, CBCA, Criteria Based Content Analysis, credibility, adult, statement, allegation, deception detection, lie detection, truthful account, Statement Validity Assessment, SVA*. Con estos descriptores se construyeron algoritmos de búsqueda que se aplicaron en todas las búsquedas electrónicas.

Criterios de inclusión y exclusión

Aunque los criterios de realidad se aplican generalmente en el contexto judicial para dotar de valor de prueba al testimonio del denunciante, observamos que la literatura también lo ha aplicado a testigos y agresores por lo que tomamos también estas poblaciones al ser el número de estudios suficientes para realizar meta-análisis. Si bien, el concepto de adulto en el contexto judicial se asocia a los 18 años y así hemos observado que lo seguían la gran mayoría de estudios, al igual que la mayoría de edad civil, algunos estudios han tomado como adultos a participantes con 17 años. Dado que esa diferencia en la edad no afecta a la capacidad para testimoniar ni cognitiva ni legalmente, se tomaron los estudios que incluían participantes con 17 años y que los autores referían como población adulta. Para su inclusión como estudios primarios, éstos debían aportar bien el tamaño del efecto de los criterios de realidad analizados para la discriminación entre declaraciones reales y fabricadas, bien datos estadísticos que posibilitaran su cómputo, incluyendo aquellos con errores en el análisis de datos pero de los que se pudo computar aun así el tamaño del efecto.

Se excluyeron aquellos datos en los que la unidad de análisis no era la declaración o que unían criterios del CBCA en un nuevo criterio (incumplimiento del requerimiento 'exclusión mutua' para crear sistemas categoriales metodícos). En los criterios adicionales se excluyeron aquellos que

no se formularon como adicionales al CBCA o que sólo eran específicos de un contexto. Igualmente, se han eliminado los datos publicados duplicadamente o más (se tomó sólo uno de ellos), pero no la fragmentación de datos de un estudio (datos independientes).

Finalmente se han seleccionado 39 estudios primarios que cumplen los criterios de inclusión y exclusión. Para el total del CBCA se calcularon 31 tamaños del efecto, mientras que para los criterios individuales, el rango oscilaba entre 5 tamaños del efecto para los criterios 10 y 19, y 35 para los criterios 3 y 8.

Procedimiento

Se siguieron las etapas establecidas por Botella y Gambara (2006) para el proceso de realización de un meta-análisis. Una vez llevada a cabo la búsqueda en la literatura e identificado aquellos artículos que formaron parte del presente meta-análisis, se procedió a la codificación de los mismos en aquellas variables que pudieran ejercer un rol moderador. A este respecto, la literatura precedente (Fariña, Arce y Real, 1994; Höfer et al., 1993; Raskin et al., 1991; Volbert y Steller, 2014; Vrij, 2005), meta-análisis previos con población de menores (Amado et al., 2015), el paradigma de investigación (campo vs. estudios experimentales) bajo la jurisprudencia legal estadounidense (Daubert v. Merrell Dow Pharmaceuticals, 1993), el cumplimiento del criterio de publicación de los Daubert estándar (DSPC, por iniciales en inglés) (revistas con revisión por pares) para ser considerada evidencia científica admisible como evidencia legal, el entrenamiento de los participantes en la condición de mentira en los criterios de realidad (condición de mentira entrenada) y la versión del sistema categorial (todos los criterios de realidad vs. versión de 14 criterios). Siguiendo un procedimiento de aproximaciones sucesivas (Fariña, Arce y Novo, 2002), en la codificación de los estudios primarios se detectaron los siguientes moderadores: estatus del declarante (víctima, agresor y testigo), tipo de evento (eventos auto-experimentados y eventos observados en vídeo) y el contexto judicial, esto es, tipo de caso.

Dado que se han formulado criterios adicionales y algunos fueron renombrados por los investigadores, cuando la denominación no era, tal cual, la original (Steller y Köhnken, 1989), se procedió por medio de una evaluación al *estilo Thurstone* en la que 10 jueces evaluaban el grado de solapamiento entre la categoría del estudio y la original. Cuando el intervalo entre Q1 y Q3 entraba en la región de independencia de criterios se tomaba como un criterio adicional, en tanto cuando entraba en la región de dependencia con el original, se tomaba como criterio original.

La codificación de los estudios y moderadores fue llevado a cabo por dos investigadores de manera independiente que se mostraron totalmente coincidentes ($\kappa=1$).

Análisis de datos

Los tamaños del efecto se tomaron directamente de los estudios primarios cuando los proporcionaban y, en su defecto, se computó el tamaño del efecto d a partir de las medias y desviaciones típicas/error estándar de la media (Cohen's d cuando $N_1=N_2$ y Glass's Δ cuando $N_1 \neq N_2$), el valor

de t , o el valor de F . Cuando los resultados estaban en proporciones se obtuvo el tamaño del efecto δ (Hedges y Olkin, 1985), que se corresponde con d , en tanto que cuando se facilitaban en tablas de contingencia 2x2 se obtuvo ϕ , que fueron transformados a d de Cohen.

La realización de los meta-análisis se llevó a cabo siguiendo el procedimiento establecido por Hunter y Schmidt (2015), tomando como unidad de análisis (n), el número de declaraciones, ponderando el tamaño del efecto por el tamaño de la muestra i.e., el número de declaraciones (d_n) y corrigiendo el tamaño del efecto por la fiabilidad del criterio (δ).

Las diferencias entre tamaños del efecto se estimaron por medio de la diferencia entre correlaciones (q statistic; Cohen, 1988), transformando para ello los tamaños del efecto a correlaciones. En el estudio de moderadores se computó el promedio de los criterios para cada moderador.

Para el estudio de la utilidad en la práctica forense de los resultados de los meta-análisis se tomaron tres estadísticos más útiles al respecto (Amado et al., 2015): U1, *Binomial Effect Size Display* (BESD) y Probabilidad de Superioridad (PS).

Fiabilidad del criterio

No todos los estudios primarios aportaban información sobre la fiabilidad o acuerdo entre-jueces para los criterios de realidad y para la puntuación total en el CBCA. Además, los coeficientes de fiabilidad informados variaban de unos estudios a otros y, en algunos estudios, se presentaban varios, tomando, en este último caso, aquellos que mejor se ajustaban a los resultados obtenidos por Anson, Golding y Gully (1993), y Horowitz et al. (1997). Dada la falta de información de la fiabilidad de la codificación en estudios y criterios concretos, estimamos la fiabilidad promedio para los criterios y para la puntuación total en el CBCA, dado que no es la misma para los criterios que para el instrumento (Horowitz et al., 1997). Para dicha estimación, sólo se tomaron los coeficientes de fiabilidad, ya que los índices de acuerdo no miden fiabilidad. Así, hallamos a partir de 172 coeficientes de fiabilidad de los criterios del CBCA en los estudios primarios una fiabilidad promedio para los criterios de realidad del CBCA de $r=0,61$ ($EEM=0,020$, $95\%CI=0,57$, $0,65$); y para la puntuación total del CBCA se procedió con la Spearman-Brown obteniéndose una $r=0,97$. Asimismo, se calculó la fiabilidad promedio para los criterios adicionales de realidad, hallando, a partir de 7 coeficientes de fiabilidad, una r de $0,74$ ($EEM=0,041$, $95\%CI=0,66$, $0,82$). A veces, se ha considerado una debilidad del sistema si la fiabilidad observada promedio es baja. Sin embargo, esta potencial deficiencia metodológica se corrige por la falta de fiabilidad del criterio en el procedimiento de meta-análisis de Hunter y Schmidt (2015).

Resultados

Estudio de outliers

Se procedió a una depuración inicial de outliers para cada uno de los criterios de realidad, la puntuación total en el CBCA y condiciones. Como criterio (casos extremos) se eligió el $\pm 3 \times IQR$ (rango inter-cuartil) del tamaño del efecto

promedio ponderado por el tamaño de la muestra, dado que los resultados de criterios más conservadores, tal como $\pm 1.5 \times IQR$ o $\pm 2SD$, eliminaban más del 10% de los tamaños del efecto por lo que es probable que no se trate de outliers, sino de moderadores (Tukey, 1960).

Meta-análisis global de los criterios de realidad en adultos

Los resultados (Tabla 2) muestran un tamaño del efecto verdadero (δ) positivo (entre presencia de criterios y realidad de la declaración), y significativo (cuando el intervalo de confianza no incluye el valor cero, indica que el tamaño del efecto es significativo) para los criterios de realidad del CBCA, a excepción de los criterios 17 'auto-desaprobación' y 18 'perdón al autor del delito', así como para la puntuación total en el CBCA. No obstante, estos resultados no son generalizables (los criterios 10 y 19 están afectados por el 'error de muestreo de segundo orden' por lo que sus resultados no son válidos para esta estimación) a futuras muestras (cuando el intervalo de credibilidad incluye el cero, indica que el tamaño del efecto no es generalizable al 80% de otras muestras). En los criterios adicionales (Höfer et al., 1993), los meta-análisis mostraron un tamaño del efecto corregido positivo, significativo y generalizable para los criterios 'estilo de la declaración' y 'dar muestras de inseguridad'. El tamaño del efecto para el criterio 'repeticiones' fue *negativo* y significativo, pero no generalizable, confirmando que no se trata de un criterio de realidad. Para los criterios 'dar explicaciones de la falta de memoria' y 'clichés', el tamaño del efecto obtenido no fue significativo. Los criterios repeticiones y clichés, que fueron formulados para discriminar entre declaraciones de hechos auto-experimentados y fabricados, se relacionan con la fabricación de hechos, es decir, no son criterios de realidad en sí por lo que no serán incluidos en análisis posteriores. La regla del 75% y el intervalo de credibilidad (Hunter y Schmidt, 2015) aconsejan el estudio de moderadores.

Estudio de moderadores

El estudio de los moderadores (se tomó el promedio de criterios como variable dependiente; Tabla 3) pone de manifiesto un tamaño del efecto positivo y significativo, pero no generalizable, en todos los moderadores analizados. Respecto a la magnitud de los tamaños del efecto, excluyendo la condición de testigo (de un evento observado en vídeo) con un tamaño del efecto moderado ($\delta > 0,50$), todos fueron pequeños ($0,20 > \delta < 0,50$). A este respecto, Arce y Fariña (2009) han sugerido (y creado) la concreción de sistemas categoriales basándose en un procedimiento *bottom-up* y no *top-down*, de modo que sólo las categorías realmente discriminativas entre memorias de hechos vividos y fabricados formen parte del sistema. De este modo se maximiza la eficacia del sistema categorial resultante, eliminando el ruido introducido por aquellas categorías creadas *top-down* no discriminativas. En consecuencia, hemos repetido los meta-análisis con las categorías de análisis de contenido cuyo tamaño del efecto era significativo i.e., el intervalo de confianza para d no contiene el cero. Los resultados (Tabla 3) revelan un incremento significativo en el

Tabla 2 Resultados del Meta-análisis Global para los Criterios de Realidad, Puntuación Total en el CBCA y Criterios Adicionales.

Criterios CBCA	k	n	d_w	SD_d	SD_{pre}	SD_{res}	δ	SD_δ	%Var	95% CI_δ	80% CV_δ
1. Estructura lógica	30	2.265	0,48	0,6990	0,2503	0,6527	0,62	0,8493	13	0,40, 0,56	-0,46, 1,71
2. Producción inestructurada	27	1.987	0,53	0,9241	0,2570	0,8876	0,69	1,1551	8	0,45, 0,61	-0,79, 2,17
3. Cantidad de detalles	35	2.714	0,55	0,8294	0,2529	0,7899	0,71	1,0279	9	0,47, 0,63	-0,60, 2,03
4. Engranaje contextual	29	2.137	0,19	0,6169	0,2372	0,5868	0,24	0,7411	15	0,11, 0,27	-0,70, 1,19
5. Descripción de interacciones	29	2.243	0,27	0,3742	0,2349	0,2912	0,36	0,3790	39	0,19, 0,35	-0,13, 0,84
6. Reproducción conversaciones	34	2.528	0,34	0,4990	0,1780	0,4662	0,44	0,6067	13	0,26, 0,42	-0,33, 1,22
7. Complicaciones inesperadas	29	1.956	0,25	0,3788	0,2498	0,2847	0,32	0,3705	43	0,17, 0,33	-0,15, 0,79
8. Detalles inusuales	35	2.441	0,31	0,6532	0,2489	0,6039	0,41	0,7859	14	0,23, 0,39	-0,59, 1,42
9. Detalles superfluos	27	1.863	0,14	0,5676	0,2437	0,5126	0,18	0,6670	18	0,04, 0,24	-0,67, 1,04
10. Detalles incomprensidos	5	376	0,22	0,1208	0,2357	0,0000	0,28	0,0000	100	0,02, 0,42	0,28
11. Asociaciones externas	22	1.612	0,26	0,4781	0,2405	0,3268	0,34	0,5376	25	0,16, 0,36	-0,35, 1,02
12. Estado mental subjetivo	28	2.170	0,18	0,4843	0,2312	0,4256	0,23	0,5538	23	0,10, 0,26	-0,47, 0,94
13. Estado mental agresor	31	2.232	0,09	0,6212	0,2376	0,5741	0,11	0,7470	15	0,01, 0,17	-0,84, 1,07
14. Correcciones espontáneas	29	1.842	0,16	0,5276	0,2545	0,4622	0,20	0,6014	23	0,06, 0,26	-0,56, 0,97
15. Admisión falta de memoria	34	2.305	0,25	0,3823	0,2494	0,2897	0,32	0,3770	42	0,17, 0,33	-0,16, 0,80
16. Dudas propio testimonio	26	1.755	0,20	0,4521	0,2478	0,3781	0,26	0,4919	30	0,10, 0,30	-0,37, 0,89
17. Auto-desaprobación	13	948	0,04	0,4629	0,2354	0,3985	0,05	0,5186	26	-0,08, 0,16	-0,61, 0,71
18. Perdón al autor del delito	8	680	-0,02	0,2796	0,2178	0,1753	-0,02	0,2281	61	-0,18, 0,14	-0,31, 0,27
19. Detalles agresión	5	562	0,28	0,1894	0,1966	0,0000	0,36	0,0000	100	0,12, 0,44	0,36
PUNTUACIÓN TOTAL CBCA	31	2.124	0,55	0,6759	0,2475	0,6290	0,56	0,6386	13	0,47, 0,63	-0,25, 1,37
PROMEDIO (criterios originales)	46	3.223	0,25	0,5032	0,2368	0,4269	0,33	0,5614	32	0,17, 0,33	-0,39, 1,05
<i>Criterios Adicionales</i>											
Estilo declaración	3	357	0,41	0,2030	0,1874	0,0781	0,48	0,0909	85	0,20, 0,63	0,36, 0,59
Dar muestras inseguridad	3	297	0,67	0,5540	0,2111	0,5122	0,78	0,5965	14	0,43, 0,90	0,01, 1,54
Dar explicaciones falta memoria	4	447	0,15	0,2877	0,1902	0,2158	0,18	0,2514	44	-0,03, 0,33	-0,14, 0,50
Clichés	3	267	-0,18	0,5145	0,2134	0,4682	-0,21	0,5452	17	-0,41, 0,05	-0,90, 0,49
Repeticiones	4	417	-0,47	0,5851	0,2011	0,5494	-0,54	0,6399	12	-0,67, -0,27	-1,36, 0,27
PROMEDIO (originales+adicionales)	46	3.223	0,27	0,4821	0,2313	0,4053	0,34	0,5275	34	0,19, 0,35	-0,33, 1,01

Nota. k = número de estudios; n = tamaño total de la muestra; d_w = tamaño del efecto ponderado por el tamaño de la muestra; SD_d = desviación estándar observada de d; SD_{pre} = desviación estándar observada de d corregida por todos los artefactos; SD_{res} = desviación estándar observada de d después de eliminar la varianza debida a todos los artefactos; δ = tamaño del efecto corregido por la fiabilidad del criterio; SD_δ = desviación estándar de δ ; %Var = varianza explicada por todos los errores artificiales; 95% CI_δ = 95% intervalo de confianza para d; 80% CV_δ = 80% intervalo de credibilidad para δ .

Tabla 3 Resultados de los Meta-análisis de Moderadores.

MODERADOR	k	n	d_w	SD_d	SD_{pre}	SD_{res}	δ	SD_δ	%Var	95% CI _d	80% CV _δ
CBCA criterios significat.(17)	46	3.223	0,27	0,5187	0,2380	0,4433	0,36	0,5835	31	0,19, 0,35	-0,39, 1,11
Versión 14 criterios	45	3.143	0,28	0,5567	0,2394	0,4906	0,36	0,6465	25	0,22, 0,34	-0,47, 1,19
<i>Criterio de publicación de los Daubert estándar</i>											
Todos los criterios (22)	35	2.256	0,20	0,4575	0,2407	0,3733	0,26	0,4786	39	0,12, 0,28	-0,35, 0,87
<i>Eventos auto-experimentados</i>											
Todos los criterios (22)	34	2.277	0,26	0,4647	0,2371	0,3879	0,33	0,5022	40	0,18, 0,34	-0,31, 0,97
<i>Eventos no auto-experimentados (testigo)</i>											
Todos los criterios (13)	11	625	0,39	0,5835	0,2707	0,5032	0,51	0,6548	65	0,23, 0,55	-0,33, 1,35
<i>Agresores</i>											
Todos los criterios (21)	11	1.067	0,27	0,4662	0,2024	0,3743	0,35	0,4975	41	0,15, 0,39	-0,29, 0,99
<i>Víctimas</i>											
Todos los criterios (18)	11	840	0,27	0,4781	0,2355	0,4012	0,35	0,5221	35	0,13, 0,41	-0,32, 1,02
<i>Estudios de campo</i>											
Todos los estudios (18)	6	422	0,34	0,4948	0,2385	0,4153	0,45	0,5404	35	0,14, 0,54	-0,24, 1,14
Criterios significativos (10) ¹	6	422	0,53	0,4774	0,2458	0,3834	0,69	0,4989	42	0,33, 0,73	0,05, 1,33
<i>Estudios de campo de violencia sexual y de género</i>											
Todos los criterios (17) ²	5	263	0,67	0,3587	0,2871	0,1957	0,87	0,2459	72	0,41, 0,92	0,55, 1,18
Criterios significativos (15) ³	5	263	0,74	0,3654	0,2892	0,2134	0,96	0,2478	72	0,48, 0,99	0,64, 1,28
<i>Estudios experimentales</i>											
Todos los criterios (22)	39	2.721	0,25	0,4497	0,2336	0,3934	0,32	0,4933	37	0,17, 0,33	-0,31, 0,95

Nota. ¹criterios significativos (criterios del CBCA): 1-3, 5-8, 11, 12 y 19; ²criterios significativos (criterios del CBCA): 1-9, 11-18; ³criterios significativos (criterios del CBCA): 1-9, 11-12, 14-17. Para los criterios adicionales, los estudios fueron insuficientes.

tamaño del efecto en los estudios de campo, $q_c=0,119$, $p<0,05$ (unidireccional; se espera un tamaño del efecto mayor con los criterios significativos), de modo que el tamaño del efecto es significativamente mayor con los criterios significativos. Además, con estos, los resultados pasan (con todos los criterios de realidad no eran generalizables) a ser generalizables (el intervalo de credibilidad no incluye el cero). En los estudios experimentales que incluyen a los restantes moderadores, los resultados no contrastan esta hipótesis pues las categorías de realidad eran previa o posteriormente depuradas eliminando las no significativas.

La técnica meta-analítica no tiene en cuenta la solidez teórica o fidelidad de los estudios incluidos con las teorías originales, es decir, incluye todos los estudios que se acogen en su objetivo al estudio de categorías de realidad. Por otra parte, los diseños experimentales empleados en estudios de testigos no son, de facto, testigos de hechos realmente auto-experimentados, sino de eventos no auto-experimentados (esto es, observados en vídeos, no implicando hechos vividos) que no cumplen con la teoría original que hipotetiza que los criterios de realidad discriminan entre memorias de hechos auto-experimentados en la vida real y hechos fabricados o ficticios. Sólo en uno de los estudios de testigos el evento era auto-experimentado, testigo presencial (Gödert, Gamer, Rill y Vossel, 2005), hallando en el total de criterios de realidad que discriminaba significativamente entre las memorias de los testigos presenciales y de agresores reales que dan un falso testimonio, $d=0,59$, $1-\beta=0,78$, y de participantes no implicados en el evento, $d=0,83$, $1-$

$\beta=0,96$. No obstante, los criterios de realidad también discriminan entre memorias de eventos observados en vídeo y eventos fabricados. El único estudio (Lee, Klaver y Hart, 2008) que compara memorias de hechos auto-experimentados (condición de realidad) y hechos observados en vídeo (condición de mentira) halló que criterios de realidad del CBCA, así como la puntuación total en el CBCA, diferenciaban significativamente entre ambas memorias en línea con la hipótesis Undeutsch.

La elevada variabilidad observada en los tamaños del efecto en los estudios de campo, que era debida fundamentalmente a un único estudio, nos llevó a buscar la causa a través de las diferencias en los diseños (el contexto delictivo en este estudio era diferente al de los restantes estudios). Como quiera el efecto de contexto se ha hipotetizado (Köhnken, 1996; Volbert y Steller, 2014) y hallado (Arce, Fariña y Vilariño, 2010; Vilariño, Novo y Seijo, 2011) que media la eficacia discriminativa de las categorías de realidad, replicamos los meta-análisis de estudios de campo en casos de violencia sexual y de género (delitos cometidos en el ámbito privado según la categorización de Arce y Fariña, 2005). Los resultados exhibieron un tamaño del efecto positivo, significativo y generalizable (no generalizable tomando todos los estudios de campo) para las investigaciones bajo esta condición. Además, la magnitud de los tamaños del efecto incrementaron significativamente tanto comparando la eficacia discriminativa de todos los criterios de realidad (0,45 para todos los estudios de campo vs. 0,87 para los casos de violencia sexual y de género,

$q_c=0,199$, $p<0,01$ (unidireccional; se espera un tamaño del efecto mayor en contextos delictivos específicos), como limitada a los criterios significativos, $q_c=0,168$, $p<0,05$ (0,69 vs. 0,96). Asimismo, los criterios de realidad fueron significativamente más útiles, $q_c=0,262$, $p<0,01$, en casos de violencia sexual y de género que en todo tipo de casos (0,32 vs. 0,87).

Los resultados (no pudo ejecutarse un meta-análisis porque *ks* y *ns* eran insuficientes y los diseños de investigación no comparables) de la comparación entre declaraciones de participantes instruidos para mentir (condición de mentira entrenada) con declaraciones verdaderas no son concluyentes¹ sobre la efectividad de los criterios de realidad para discriminar entre declaraciones verdaderas y falsas.

Discusión

De los anteriores resultados se extraen las siguientes conclusiones. Primera, los resultados confirman la hipótesis Undeutsch, es decir, los criterios de realidad discriminan entre memorias de hechos auto-experimentados y fabricados [el sesgo de publicación, File Drawer Analysis (FDA; Hunter y Schmidt, 2015) mostró que para reducir esta hipótesis a un efecto trivial (McNatt, 2000), 0,05, en el promedio de los criterios del CBCA, se necesitarían 184 estudios con un efecto nulo, contingencia que es muy improbable que se dé]. Esta hipótesis, que cumple con el DSPC es, además, válida tanto para memorias de víctimas/denunciantes y agresores (se necesita más investigación para memorias de eventos auto-experimentados en testigos presenciales); y robusta tanto en estudios experimentales (alta validez interna) como de campo (alta validez externa). No obstante, los criterios de realidad también discriminan entre memorias de eventos observados en vídeo, esto es, no auto-experimentados, y eventos fabricados, contingencias para las que no ha sido formulada la hipótesis. Además, la investigación no es concluyente respecto a la validez de la hipótesis con sujetos entrenados para mentir. Segunda, asimismo, los resultados también validan el CBCA, como sistema categorial basado en la hipótesis Undeutsch. Ahora bien, no todos los criterios son válidos, al tiempo que no son generalizables e incluso pueden ser contrarios a la hipótesis. En consecuencia, dichos criterios no se pueden utilizar en todo tipo de contextos, ni indiscriminadamente. Ambas versiones del CBCA (todos los criterios y 14 criterios) son exactamente iguales ($\delta=0,36$) a la hora de discriminar entre hechos auto-experimentados y fabricados. Además, los resultados abren la puerta a la inclusión de nuevos criterios de realidad, pero también advierten que se han postulado criterios adicionales que no cumplen la hipótesis Undeutsch (tamaños del efecto significativamente negativos, es decir, criterios de no realidad) por lo que no pueden ser incluidos en el CBCA. Tercera, en los estudios de campo encontramos

que la capacidad discriminativa de los criterios de realidad se incrementa de modo significativo en casos de violencia sexual y de género (FDA: para reducir los resultados en casos de violencia sexual y de género a un efecto trivial, serían necesarios 62 y 69 estudios con un efecto nulo, para todos los criterios y para los significativos, respectivamente). Es muy improbable que esto ocurra frente a otro tipo de contextos (FDA: para reducir la eficacia discriminativa de los criterios de realidad entre memorias de hechos reales y fabricados en estudios de campo y en cualquier contexto a un efecto trivial, se necesitarían 35 estudios con un efecto nulo, siendo improbable que esto suceda). Sucintamente, las áreas de ambas poblaciones no se superponen en el 54% ($U1=0,54$), es decir, son totalmente independientes y, por tanto, la eficacia discriminativa de los criterios de realidad entre memorias de eventos auto-experimentados y fabricados en casos de violencia sexual y de género es total en el 54% de las evaluaciones de credibilidad. Además, el 75% de las declaraciones de hechos auto-experimentados contienen más criterios de realidad que las fabricadas (probabilidad de superioridad, $PS=0,75$), siendo la probabilidad de falsos positivos del 28% (BESD). Estos resultados son altamente robustos pues no sólo establecen una relación positiva y significativa entre criterios de realidad y declaraciones verdaderas, sino que son generalizables a todo tipo de casos de violencia sexual y de género, y son homogéneos (es decir, sujetos a escasa variabilidad dado que la correlación entre los tamaños del efecto es de 0,72).

En relación con las implicaciones prácticas, de esta revisión se deduce que los criterios de realidad de las declaraciones son estadísticamente efectivos en la discriminación entre memorias de hechos auto-experimentados y fabricados, pero esto no implica que sean directamente generalizables para la práctica forense. Así, en las mejores condiciones de discriminación, estudios de campo en casos de violencia sexual y de género, la probabilidad de falsos positivos puede alcanzar 0,22, cuando en el ámbito forense dicha probabilidad ha de ser cero (Arce, Fariña y Fraga, 2000). En general, en la práctica forense sólo se pueden tomar los criterios de realidad significativos, esto es, avalados por evidencia científica (véase la nota al pie en Tabla 3) pues los resultados son generalizables, en tanto que con el total de criterios no lo son. Ahora bien, dado que el límite inferior del intervalo de credibilidad fue 0,05, la utilidad práctica de dichas categorías puede ser prácticamente despreciable ($PS=0,51$), es decir, únicamente en el 51% de las declaraciones verdaderas habría más criterios de realidad que en las falsas, desconociéndose bajo qué condiciones sería esta contingencia. Sin embargo, el límite inferior del intervalo de credibilidad de los criterios de realidad aplicados a casos de violencia sexual y de género, que también son generalizables, tanto con todos los criterios como con los significativos, es mayor, $PS=0,73$ y $.75$ (Hedges y Olkin's $\delta=0,59$ y $0,65$, valor de prueba=0,51), para todos los criterios de realidad y para los criterios significativos, respectivamente. Pero, estas conclusiones no son de aplicación directa en la práctica forense porque se desconoce cuál es el criterio de decisión que en el contexto forense ha de ser un 'criterio de decisión estricto' en el que el error tipo II (clasificación de una declaración falsa como verdadera) no es admisible, esto es, ha de ser igual a cero. Sobre esto, Arce

¹ Las conclusiones de efectos no significativos en los estudios primarios no son concluyentes, ya que la potencia estadística de los diseños, $1-B<0,80$, es insuficiente para esta conclusión ($d=-0,44$, $1-B=0,41$, Bogaard, Meijer y Vrij, 2013; $d=0,37$, $1-B=0,26$, Vrij, Akehurst, Soukara y Bull, 2002; $d=0,11$, $1-B=0,06$, Vrij, Kneller y Mann, 2000).

et al. (2010) encontraron en declaraciones fabricadas en casos de violencia de género hasta 13 criterios de realidad del CBCA por lo que se precisaría detectar, al menos, 14 criterios de realidad en la declaración para concluir que es verdadera, con una correcta clasificación de verdaderos positivos (declaraciones reales clasificadas como tales) del 36%. En suma, los criterios de realidad del CBCA son una herramienta poco eficaz para dotar de valor de prueba al testimonio de la denunciante de violencia de género. Por ello, han de ser complementados con criterios adicionales que los doten de mayor eficacia. Complementando los criterios de realidad del CBCA, con criterios del SRA, atributos de memoria, y criterios de realidad adicionales propios de casos de violencia de género creados a partir de declaraciones reales (sentencia judicial como *ground truth*). Arce y Fariña (2009), Vilariño (2010) y Vilariño et al. (2011) crearon y validaron un sistema categorial específico de casos de violencia de género, incluyendo violencia sexual, con el que con un criterio de decisión estricto lograron reducir la tasa de falsos negativos al 2%. En todo caso, sólo son transferibles a la práctica forense aquellos resultados que ofrezcan un criterio de decisión estricto y con una clasificación total de los falsos positivos (tasa=0).

Para la futura investigación, los resultados del presente meta-análisis advierten de la necesidad de más investigación con diseños experimentales orientados a conocer si los criterios de realidad diferencian entre memorias de eventos auto-experimentados y eventos observados en vídeo pero no experimentados; entre eventos vividos como testigo presencial y eventos fabricados; entre memorias de participantes entrenados para mentir y de participantes honestos; así como investigaciones que conduzcan a la detección de nuevas categorías de realidad (*bottom-up*), principalmente para contextos específicos, esto es, victimización de delitos concretos.

Este meta-análisis está sujeto a las siguientes limitaciones. Primera, la investigación publicada ha sesgado los resultados de modo que los resultados no significativos o de categorías previsiblemente ineficaces fueron eliminados (favorece la conformación de la hipótesis Undeutsch). Segunda, la metodología de simulación (estudios experimentales) no tiene probada validez externa (Sarwar, Allwood y Innes-Ker, 2014), sino *face validity* (Konecni y Ebbesen, 1992). Tercera, en parte de la literatura experimental las declaraciones constituyen material insuficiente para la evaluación de la realidad del testimonio aplicando análisis de contenido (Köhnken, 2004), favoreciendo el rechazo de la hipótesis Undeutsch. Cuarta, no hay un control de los efectos del entrevistador en los contenidos de la declaración, desconociéndose la fiabilidad de las entrevistas siendo éstas realizadas, además, por entrevistadores poco entrenados. Quinta, no hay constancia en muchos estudios de que se siguieran los estándares del SVA que constituyen un requerimiento para la aplicación del CBCA. Sexta, los resultados de algunos meta-análisis pueden estar sujetos a cierta variabilidad, tratándose de $N_s < 400$, no garantizan la estabilidad de las estimaciones muestrales (Hunter y Schmidt, 2015). Séptima, algunos estudios primarios no estimaron la fiabilidad de las codificaciones, por lo que la fiabilidad de estos resultados es una estimación.

Financiación

Esta investigación ha sido financiada en los Proyectos de I+D+I, del Programa Estatal De Investigación, Desarrollo e Innovación Orientada a los Retos de la Sociedad del Ministerio de Economía y Competitividad del Gobierno de España al proyecto referencia PSI2014-53085-R.

Referencias

- *Akehurst, L., Easton, S., Fullar, E., Drane, G., Kuzmin, K. y Litchfield, S. (2015). An evaluation of a new tool to aid judgements of credibility in the medico-legal setting. *Legal and Criminological Psychology*. Advance online publication. doi:10.1111/lcrp.12079
- Amado, B.G., Arce, R. y Fariña, F. (2015). Undeutsch hypothesis and Criteria Based Content Analysis: A meta-analytic review. *European Journal of Psychology Applied to Legal Context*, 7, 3-12. doi:10.1016/j.ejpal.2014.11.002
- Anson, D.A., Golding, S.L. y Gully, K.J. (1993). Child sexual abuse allegations: Reliability of Criteria-Based Content Analysis. *Law and Human Behavior*, 17, 331-341. doi:10.1007/BF01044512
- Arce, R. y Fariña, F. (2005). Peritación psicológica de la credibilidad del testimonio, la huella psíquica y la simulación: El Sistema de Evaluación Global (SEG). *Papeles del Psicólogo*, 26, 59-77.
- Arce, R. y Fariña, F. (2009). Evaluación psicológica forense de la credibilidad y daño psíquico en casos de violencia de género mediante el Sistema de Evaluación Global. En F. Fariña, R. Arce y G. Buela-Casal (Eds.), *Violencia de género. Tratado psicológico y legal* (pp. 147-168). Madrid: Biblioteca Nueva.
- Arce, R. y Fariña, F. (2012). Psicología social aplicada al ámbito jurídico. En A.V. Arias, J.F. Morales, E. Nouvilas y J.L. Martínez (Eds.), *Psicología social aplicada* (pp. 157-182). Madrid: Panamericana.
- Arce, R., Fariña, F. y Fraga, A. (2000). Género y formación de juicios en un caso de violación. *Psicothema*, 12, 623-628.
- *Arce, R., Fariña, F. y Vilariño, M. (2010). Contraste de la efectividad del CBCA en la evaluación de la credibilidad en casos de violencia de género. *Intervención Psicosocial*, 19, 109-119. doi:10.5093/in2010v19n2a2
- *Beaulieu-Prévost, D. (2001). *Analyse de validité de la déclaration (SVA), mensonge et faux souvenirs: Validité et efficacité chez les adultes*. (Doctoral dissertation). Retrieved from ProQuest Dissertations y Theses Global. (Order No. MQ60609)
- *Bensi, L., Gambetti, E., Nori, R. y Giusberti, F. (2009). Discerning truth from deception: The sincere witness profile. *European Journal of Psychology Applied to Legal Context*, 1, 101-121.
- Berliner, L. y Conte, J.R. (1993). Sexual abuse evaluation: Conceptual and empirical obstacles. *Child Abuse and Neglect*, 17, 111-125. doi:10.1016/0145-2134(93)90012-T
- *Biland, C., Py, J. y Rimboud, S. (1999). Evaluer la sincérité d'un témoin grâce à trois techniques d'analyse, verbales et non verbales. *European Review of Applied Psychology-Revue Européenne de Psychologie Appliquée*, 49, 115-122.
- *Blandón-Gitlin, I., Pezdek, K., Lindsay, D.S. y Hagen, L. (2009). Criteria-Based Content Analysis of true and suggested accounts of events. *Applied Cognitive Psychology*, 23, 901-917. doi:10.1002/acp.1504
- *Bogaard, G., Meijer, E.H. y Vrij, A. (2013). Using an example statement increases information but does not increase accuracy of CBCA, RM, and SCAN. *Journal of Investigative Psychology and Offender Profiling*, 11, 151-163. doi:10.1002/jip.1409
- Botella, J. y Gambara, H. (2006). Doing and reporting a meta-analysis. *International Journal of Clinical y Health Psychology*, 6, 425-440.

- *Caso, L., Vrij, A., Mann, S. y de Leo, G. (2006). Deceptive responses: The impact of verbal and non-verbal countermeasures. *Legal and Criminological Psychology*, 11, 99–111. doi:10.1348/135532505X49936
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2ª ed.). Hillsdale, NJ: LEA.
- *Critchlow, N. (2011). *Applying Criteria Based Content Analysis to assessing the veracity of rape statements* (Unpublished doctoral dissertation). Manchester Metropolitan University, Manchester, UK.
- *Critchlow, N. (2011). [A field validation of CBCA when assessing authentic police rape statements: evidence for discriminant validity to prescribe veracity to adult narrative]. Unpublished raw data.
- *Dana-Kirby, L. (1997). *Discerning truth from deception: Is Criteria-Based Content Analysis effective with adult statements?* (Unpublished doctoral dissertation). University of Oregon, Oregon.
- Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993).
- *Evans, J., Michael, S.W., Meissner, C.A. y Brandon, S.E. (2013). Validating a new assessment method for deception detection: Introducing a psychologically based credibility assessment tool. *Journal of Applied Research in Memory and Cognition*, 2, 33–41. doi:10.1016/j.jarmac.2013.02.002
- Fariña, F., Arce, R. y Novo, M. (2002). Heurístico de anclaje en las decisiones judiciales. *Psicothema*, 14, 39–46.
- Fariña, F., Arce, R. y Real, S. (1994). Ruedas de identificación: De la simulación y la realidad. *Psicothema*, 6, 395–402.
- *Gödert, H.W., Gamer, M., Rill, H.G. y Vossel, G. (2005). Statement Validity Assessment: Inter-rater reliability of Criteria-Based Content Analysis in the mock-crime paradigm. *Legal and Criminological Psychology*, 10, 225–245. doi:10.1348/135532505X52680
- *Godoy, V. y Higuera, L. (2008). El análisis de contenido basado en criterios (CBCA) y la entrevista cognitiva aplicados a la credibilidad del testimonio en adultos. En F. Rodríguez, C. Bringas, F. Fariña, R. Arce y A. Bernardo (Eds.), *Psicología Jurídica: Entorno judicial y delincuencia* (pp. 117–125). Recuperado de <http://gip.uniovi.es/T5EJD.pdf>
- Hans, V.P. y Vidmar, N. (1986). *Judging the jury*. Nueva York: Plenum Press.
- Hedges, L.V. y Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press
- Höfer, E., Köhnken, G., Hanewinkel, R. y Bruhn, C. (1993). *Diagnostik und attribution von glaubwürdigkeit*. Unpublished final report. University of Kiel, Germany.
- *Honts, C.R. y Devitt, M.K. (1993). *Credibility Assessment of Verbatim Statements (CAVS)*. Retrieved from <http://www.dtic.mil/dtic/tr/fulltext/u2/a271575.pdf>
- Horowitz, S.W., Lamb, M.E., Esplin, P.W., Boychuk, T.D., Krispin, O. y Reiter-Lavery, L. (1997). Reliability of criteria-based content analysis of child witness statements. *Legal and Criminological Psychology*, 2, 11–21. doi:10.1111/j.2044-8333.1997.tb00329.x
- Hunter, J.E. y Schmidt, F.L. (2015). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.
- *Johnston, S., Candelier, A., Powers-Green, D. y Rahmani, S. (2014). Attributes of truthful versus deceitful statements in the evaluation of accused child molesters. *Sage Open*, 4(3), 1–10. doi:10.1177/2158244014548849
- *Juárez, J.R., Mateu, A. y Sala, E. (2007). *Criterios de evaluación de la credibilidad en las denuncias de violencia de género*. Recuperado de <http://justicia.gencat.cat/web/.content/document/arxiu/sc-3-143-07-cas.pdf>
- Köhnken, G. (1996). Social psychology and the law. En G.R. Semin y K. Fiedler (Eds.), *Applied social psychology* (pp. 257–282). Thousand Oaks, CA: Sage.
- Köhnken, G. (2004). Statement Validity Analysis and the detection of the truth. En P.A. Granhag y L.A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 41–63). Cambridge, Cambridge University Press.
- *Köhnken, G., Schimossek, E., Aschermann, E. y Höfer, E. (1995). The cognitive interview and the assessment of the credibility of adults' statements. *Journal of Applied Psychology*, 80, 671–684. doi:10.1037/0021-9010.80.6.671
- Konecni, V.J. y Ebbesen, E.B. (1992). Methodological issues on legal decision-making, with special reference to experimental simulations. En F. Lösel, D. Bender, y T. Bliesener (Eds.), *Psychology and law. International perspectives* (pp. 413–423). Berlin: Walter de Gruyter.
- *Leal, S., Vrij, A., Warmelink, L., Vernham, Z. y Fisher, R.P. (2015). You cannot hide your telephone lies: Providing a model statement as an aid to detect deception in insurance telephone calls. *Legal and Criminological Psychology*, 20, 129–146. doi:10.1111/lcrp.12017
- *Lee, Z., Klaver, J.R. y Hart, S.D. (2008). Psychopathy and verbal indicators of deception in offenders. *Psychology, Crime and Law*, 14, 73–84. doi:10.1080/10683160701423738
- McNatt, D.B. (2000). Ancient Pygmalion joins contemporary management: A meta-analysis of the result. *Journal of Applied Psychology*, 85, 314–322. doi:10.1037/0021-9010.85.2.314
- *Merckelbach, H. (2004). Telling a good story: Fantasy proneness and the quality of fabricated memories. *Personality and Individual Differences*, 37, 1371–1382. doi:10.1016/j.paid.2004.01.007
- Novo, M. y Seijo, D. (2010). Judicial judgement-making and legal criteria of testimonial credibility. *European Journal of Psychology Applied to Legal Context*, 2, 91–115.
- *Porter, S. y Yuille, J.C. (1996). The language of deceit: An investigation of the verbal clues to deception in the interrogation context. *Law and Human Behavior*, 20, 443–458. doi:10.1007/BF01498980
- *Porter, S., Yuille, J.C. y Lehman, D.R. (1999). The nature of real, implanted, and fabricated memories for emotional childhood events: Implications for the recovered memory debate. *Law and Human Behavior*, 23, 517–537. doi:10.1023/A:1022344128649
- Raskin, D.C., Esplin, F.W. y Horowitz, S. (1991). *Investigative interviews and assessment of children in sexual abuse cases*. Manuscript no publicado, Department of Psychology, University of Utah, Utah.
- *Rassin, E. y van der Sleen, J. (2005). Characteristics of true versus false allegations of sexual offences. *Psychological Reports*, 97, 589–598. doi:10.2466/pr0.97.2.589-598
- Sarwar, F., Allwood, C.M. y Innes-Ker, A. (2014). Effects of different types of forensic information on eyewitness' memory and confidence accuracy. *European Journal of Psychology Applied to Legal Context*, 6, 17–27. doi:10.5093/ejpalc2014a3
- *Schelleman-Offermans, K. y Merckelbach, H. (2010). Fantasy proneness as a confounder of verbal lie detection tools. *Journal of Investigative Psychology and Offender Profiling*, 7, 247–260. doi:10.1002/jip.121
- *Sporer, S.L. (1997). The less travelled road to truth: Verbal cues in deception detection in accounts of fabricated and self-experienced events. *Applied Cognitive Psychology*, 11, 373–397. doi:10.1002/(SICI)1099-0720(199710)11:5<373::AID-ACP461>3.0.CO;2-0
- Steller, M. y Böhm, C. (2006). Cincuenta años de jurisprudencia del Tribunal Federal Supremo alemán sobre la psicología del testimonio. Balance y perspectiva. En T. Fabian, C. Böhm y J. Romero (Eds.), *Nuevos caminos y conceptos en la psicología jurídica* (pp. 53–67). Münster: LIT Verlag.
- Steller, M. y Köhnken, G. (1989). Criteria-Based Content Analysis. En D.C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217–245). Nueva York: Springer-Verlag.
- *Ternes, M. (2009). *Verbal credibility assessment of incarcerated violent offenders' memory reports*. Tesis Doctoral no publicada. University of British Columbia, Vancouver.

- Tukey, J.W. (1960). A survey of sampling from contaminated distributions. En I. Olkin, J.G. Ghurye, W. Hoeffding, W.G. Madow y H. Mann (Eds.), *Contributions to probability and statistics* (pp. 448–485). Stanford, CA: Stanford University Press.
- Undeutsch, U. (1967). Beurteilung der glaubhaftigkeit von aussagen. en U. Undeutsch (Ed.), *Handbuch der Psychologie, Vol. 11: Forensische Psychologie* (pp. 26–181). Göttingen: Hogrefe.
- *Vilariño, M. (2010). *¿Es posible discriminar declaraciones reales de imaginadas y huella psíquica real de simulada en casos de violencia de género?* (Tesis Doctoral, Universidad de Santiago de Compostela, Spain). Recuperado de <http://hdl.handle.net/10347/2831>
- Vilariño, M., Novo, M. y Seijo, D. (2011). Estudio de la eficacia de las categorías de realidad del testimonio del Sistema de Evaluación Global (SEG) en casos de violencia de género. *Revista Iberoamericana de Psicología y Salud*, 2, 1–26.
- Volbert, R. y Steller, M. (2014). Is this testimony truthful, fabricated, or based on false memory? Credibility assessment 25 years after Steller and Köhnken (1989). *European Psychologist*, 19, 207–220. doi:10.1027/1016-9040/a000200
- Vrij, A. (2005). Criteria-Based Content Analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law*, 11, 3–41. doi:10.1037/1076-8971.11.1.3
- Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities* (2ª ed.). Chichester: John Wiley and Sons.
- Vrij, A., Akehurst, L., Soukara, S. y Bull, R. (2002). Will the truth come out? The effect of deception, age, status, coaching and social skills on CBCA scores. *Law and Human Behavior*, 26, 261–283. doi:10.1023/A:1015313120905
- *Vrij, A., Akehurst, L., Soukara, R., y Bull, R. (2004). Detecting deceit via analyses of verbal and nonverbal behavior in children and adults. *Human Communication Research*, 30, 8–41. doi:10.1111/j.1468-2958.2004.tb00723.x
- *Vrij, A., Edward, K. y Bull, R. (2001). People's insight into their own behaviour and speech content while lying. *British Journal of Psychology*, 92, 373–389. doi:10.1348/000712601162248
- *Vrij, A., Edward, K., Roberts, K.P. y Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior*, 24, 239–263. doi:10.1023/A:1006610329284
- *Vrij, A., Evans, H., Akehurst, L. y Mann, S. (2004). Rapid judgments in assessing verbal and nonverbal cues: Their potential for deception researchers and lie detection. *Applied Cognitive Psychology*, 18, 283–296. doi:10.1002/acp.964
- *Vrij, A. y Heaven, S. (1999). Vocal and verbal indicators of deception as a function of lie complexity. *Psychology, Crime and Law*, 5, 203–215. doi:10.1080/10683169908401767
- *Vrij, A., Kneller, W. y Mann, S. (2000). The effect of informing liars about Criteria-Based Content Analysis on their ability to deceive CBCA-raters. *Legal and Criminological Psychology*, 5, 57–70. doi:10.1348/135532500167976
- *Vrij, A. y Mann, S., (2006). Criteria-Based Content Analysis: An empirical test of its underlying processes. *Psychology, Crime and Law*, 12, 337–349. doi:10.1080/10683160500129007
- *Vrij, A., Mann, S. y Edward, K. (2000). I think it was a green scarf but I am not sure. Raising doubts about one's own testimony during lying and truth telling. En A. Czerederecka, T. Jaskiewicz-Obydzinska y J. Wójcikiewicz, (Eds.), *Forensic psychology and law. Traditional questions and new ideas* (pp. 205–207). Institute of forensic research, Poland: Krakow.
- *Vrij, A., Mann, S., Kristen, S. y Fisher, R.P. (2007). Cues to deception and ability to detect lies as a function of police interview styles. *Law and Human Behavior*, 31, 449–518. doi:10.1007/s10979-006-9066-4
- *Willén, R.M. y Strömwall, L.A. (2011). Offender's uncoerced false confessions: A new application of statement analysis? *Legal and Criminological Psychology*, 17, 346–359. doi:10.1111/j.2044-8333.2011.02018.x.
- *Wojciechowski, B.W. (2014). Content analysis algorithms: An innovative and accurate approach to statement veracity assessment. *European Poligraph*, 8, 119–128. doi:0.2478/ep-2014-0010